Child Development, March/April 2004, Volume 75, Number 2, Pages 562 – 579

Executive Processes in Appearance–Reality Tasks: The Role of Inhibition of Attention and Symbolic Representation

Ellen Bialystok and Lili Senman

Two studies addressed the role of representation ability and control of attention on solutions to an appearance–reality task based on two types of objects, real and representational. In Study 1, 67 preschool children (3-, 4-, and 5-year-olds) solved appearance–reality problems and executive processing tasks. There was an interaction between object type (real vs. representational) and question type (appearance vs. reality) on problem difficulty. In addition, representational ability predicted performance on appearance questions and inhibitory control predicted performance on reality questions. In Study 2, 95 children (4- and 5-year-olds) who were monolingual or bilingual solved similar problems. On appearance questions, groups performed equivalently but on reality questions, bilinguals performed better (once language proficiency had been controlled). The difference is attributed to the advanced inhibitory control that comes with bilingualism.

One of the most significant cognitive achievements in the preschool years is the development of a theory of mind (ToM), transforming a child's notions of self and others, beliefs and facts, perceptions and understandings. Some researchers have claimed that the ability to identify false belief is the most convincing way to demonstrate this development (Astington & Gopnik, 1991a; Baron-Cohen, Leslie, & Frith, 1985; Flavell, Flavell, & Green, 1983; Whiten, 1997; Wimmer & Perner, 1983). Evidence for ToM requires demonstrating that children can ascribe beliefs to others that are different from their own (Astington & Gopnik, 1991b).

Various false-belief tasks, including unexpected location, unexpected contents, and appearance–reality (A–R), have been designed to dissociate the state of the world from the child's beliefs about that state. The paradigm was created by Flavell et al. (1983). They presented children with deceptive objects, such as a sponge that looked like a rock, and encouraged them to play with the objects to discover their properties. Using forced-choice response, children were asked what the objects looked like and what they really were. Most 3-year-olds could only think of the object in one way and committed the realist error in which they responded sponge to both questions.

The paradigm was expanded by adding a false-belief question in which children were asked what someone who had never seen the object would think it was (Astington & Gopnik, 1991a, 1991b; Frye, Zelazo, & Palfai, 1995; Gopnik & Astington, 1988). Most 4-year–olds and older children answered correctly, but 3-year-olds expected that a new observer would share their knowledge of the object and know that it was a sponge in spite of its appearance. The interpretation was that young children failed the problem for the same reason they failed the false-belief task; namely, they lacked the necessary competence with ToM to recognize that other people's minds include representations that may be different from their own (Astington & Gopnik, 1991a, 1991b). Subsequent studies have characterized A–R as ToM tasks even without the false-belief question (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002).

Across ToM tasks, there is a change at about 4 years of age when children begin to attribute knowledge properly to themselves and others. In an important meta-analysis of hundreds of studies, comprising 591 experimental conditions, Wellman, Cross, and Watson (2001) concluded that the developmental effect is robust and is only minimally modified by methodological variations in the task. Their interpretation of this impressive analysis was that it validated the coherence of the construct: "Children's performance is systematically dependent on the one thing that does not vary, namely

Ellen Bialysok and Lili Senman, Department of Psychology, York University.

The research reported in this article was funded by a grant from the Natural Sciences and Engineering Research Council of Canada to the first author. Study 1 was done in partial fulfillment of the requirements for the degree of Master of Arts by the second author. We are grateful to Jane Logan for her help in preparing the manuscript.

Correspondence concerning this article should be addressed to Ellen Bialystok, Department of Psychology, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, Canada. Electronic mail may be sent to ellenb@yorku.ca.

 0009-3920/2004/7502-0022

their conception of belief states'' (Wellman et al., 2001, p. 671).

One perspective that is consistent with this meta-analysis is the theory theory (Gopnik, 1996; Gopnik & Wellman, 1994; Perner, 1991). This approach focuses on the child's developing understanding of representational mental states, desires, intentions, perceptions, and beliefs. Children's concepts about these mental states form an integrated set of ideas, making them a unique and coherent domain of knowledge. The age-related shift in performance on ToM tasks is explained in terms of improvement in the child's conceptual competence, driven by a shift in understanding mental states. According to this view, children's development of ToM is analogous to the process of creating a scientific theory. The crucial development is the understanding of what the mind does, that is, understanding representation.

The executive functioning approach to explaining children's development of ToM is an alternative that is less compatible with the perspective advocated in the meta-analysis. According to this view, children succeed on ToM tasks as they gain control over the relevant cognitive functions implicated in these complex tasks. Proponents of this view attempt to identify those cognitive processes and to examine their predictive role in children's success. Unlike the theory theory, this approach does not consider ToM to be an autonomous area of cognitive functioning; rather, it is an extension of children's cognitive development. The two cognitive skills that have been most extensively studied under this assumption are language proficiency (Astington & Jenkins, 1999; de Villiers, 1999; de Villiers & de Villiers, 2000; Happé, 1995; Tager-Flusberg & Sullivan, 1994) and executive functioning (Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998; Frye et al., 1995; Hala, Hug, & Henderson, 2003; Hala & Russell, 2001; Hughes, 1998a, 1998b).

Executive functions are the processes that underlie goal-directed behavior, including planning, inhibition, mental flexibility, representation, and working memory (Duncan, 1986). Advances in these executive functions between the ages of 3 and 6 years coincide with the mastery of ToM, and various studies have attributed causal significance to this relationship (Carlson et al., 2002; Carlson et al., 1998; Cole & Mitchell, 2000; Davis & Pratt, 1995; Frye et al., 1995; Hala et al., 2003; Hughes, 1998a, 1998b; Perner, Lang, & Kloo, 2002; Tager-Flusberg, Sullivan, & Boshart, 1997).

One of the most frequently cited components of executive functioning is inhibition. Hughes (1998a, 1998b) studied the role of inhibitory control of motor responses in ToM tasks using Luria's (1973) hand game and her own detour-reaching task. The hand game consisted of two conditions. In the control condition, the child imitated the hand posture modeled by the experimenter (fist–fist or finger–finger), and in the conflict condition the child made the opposite posture. In the detour-reaching task, the child had to obtain a marble from a platform in a box. In some conditions, an infrared beam activated a photoelectric circuit that caused a trap door to open and the marble to disappear if the child reached directly for the marble; therefore, the child had to resist the obvious action. Both tasks were correlated with performance on the ToM tests, and the detour-reaching tasks continued to predict performance 1 year later, even when age and verbal ability were statistically controlled. Studies using other inhibition tasks have found the same effect. In the day–night Stroop task (Gerstadt, Hong, & Diamond, 1994), children are instructed to say ''day'' to a picture of a moon in a dark sky and ''night'' to a picture of a bright sun in a blue sky. They are then shown a series of these pictures and must name them as quickly and accurately as possible using this reverse rule. High performance on this task is related to success in ToM tasks (Hala et al., 2003; Tager-Flusberg et al., 1997).

The meta-analysis by Wellman et al. (2001) and the research on which it is based demonstrate the importance of the ToM concept in children's difficulties with these tasks. However, the cognitive, and particularly the executive function components of these problems are still formidable. Wellman et al. agreed that the removal of these executive function demands improves performance (cf. Carlson & Moses, 2001) but maintained that the residual difficulties that children have on these problems is attributable to a lack of a ToM. Nonetheless, given the persistence of executive processes in analyses of ToM performance, documenting their role is essential to understanding how children acquire a ToM.

The present studies investigate the role of cognitive and executive functioning in the solution to one kind of ToM problem, namely, A–R conflict tasks (Flavell et al., 1983). The solution depends on a constellation of cognitive and executive processes: Children must understand the task, represent the situation, and evaluate the information in the face of competing cues. The processing demands, therefore, include both representation of the problem and control over attention. Children must be able to choose between conflicting representations of the object without being distracted by misleading perceptual cues to provide the correct responses. The hypothesis for the first study is that children's

success will be determined in part by their ability to represent and remember the information and to inhibit attention to misleading cues. The hypothesis for the second study is that a group of children who have developed higher levels of inhibitory control will be more successful in solving these A–R problems.

Study 1

The first study examined the role of representation and inhibition in the solution to A–R problems. Representation and inhibition demands were manipulated by defining two types of objects and two types of questions. Children's abilities in representation and inhibitory control were assessed, and a simple measure of short-term verbal memory was included.

The role of object type was examined by a distinction between real objects and representational objects, determined by the relation between the object's perceptual features and its function. In real objects, the perceptual features directly indicate the function that the object is expected to serve; in representational objects, the perceptual features stand for an object that could actually serve that function. The real objects were a sponge rock (Flavell et al., 1983) and a crayons box with Lego pieces inside (adapted from Gopnik & Astington, 1988). In both cases, observers believe the object can be correctly identified by its appearance—the irregularly shaped gray properties of the sponge rock suggest a heavy rough item and the familiar packaging of the crayons box promises a set of objects that could be used for coloring. Although items such as the crayons box are sometimes labeled as unexpected contents, performance on these items is strongly correlated with that on A–R tasks, (Wellman et al., 2001) and these variants have been treated equivalently by most researchers. The representational objects were a book in the shape of a snowman and a marker pen that looks like a killer whale. In these cases, the perceptual features indicate representations of a snowman and a whale with no illusions about the reality of those objects. Children seeing these items can correctly identify their intended referents but do not mistake them for actual objects—this snowman will not melt in the sun and this whale cannot swim.

The difference between the object types is the nature of mental representation each requires. For real objects, children perceive the object's properties and represent it directly: It is a rock. For the representational objects, children interpret the perceptual properties as an indication of a familiar object: It stands for a snowman. That is, the characteristic shape and color of a snowman that is conveyed by the small plastic object is represented as something that means snowman but is not itself a snowman. Previous studies have used both real and representational objects without considering potential differences between them (Astington & Jenkins, 1999; Cole & Mitchell, 2000; Frye et al., 1995).

The second distinction was between two types of questions called the appearance question and the reality question. In A–R tasks, children are typically asked what they thought the object looked like before its identity was revealed (appearance question) and what it actually is (reality question), but their answers are combined and designate only pass or fail. However, these questions explore different aspects of children's understanding of those objects. Following Nelson (1973), mental representations are initially based on perceptual features but objects are identified by means of functions. Normally, these form–function relations are reliable: Objects that look like an x generally are an x—but in the A–R task, the associations are disrupted. This distinction between form and function corresponds to the difference between the appearance and the reality questions. The appearance question is answered on the basis of the perceptual features (it looks like an x) and the reality question on the functional properties (it is a y). If the perceptual features falsely indicate the functional identity of the object, they must be ignored to answer the reality question. Therefore, appearance questions are more directly tests of representational ability, but reality questions are more directly tests of inhibitory control. Because of these potential differences in processing, the two types of questions were examined separately.

Both distinctions indicate a bias for one of the pair to be based more on representational ability and the other on inhibitory control. For object types, representational objects require a symbolic representation in which the item stands for a meaning rather than instantiating it directly whereas real objects make no special demands on representational ability. For question type, reality questions require attentional control to ignore the salient perceptual features and state the contradictory functional identity whereas appearance questions make no special demands on inhibitory control. Therefore, representational ability and inhibitory control should differentially predict performance on real and representational objects and on appearance and reality questions. Such dissociations would help to explicate the role of executive functioning in solving A–R tasks.

Crucial to such an investigation is the method for assessing these executive skills. Diamond and her colleagues have produced several important tasks for measuring inhibitory control, including the tapping task (Diamond & Taylor, 1996) and the day–night Stroop task (Gerstadt et al., 1994). These tasks have proven to be reliable methods of gauging children's development of executive functioning.

DeLoache and colleagues have studied the development of representations and provided compelling evidence for the developmental ability to move from simple representations of objects (it is an *x*) to symbolic representations of meaning (it stands for *x*; DeLoache, 1987, 1991; DeLoache, Miller, & Rosengren, 1997). A task developed by Bialystok (1991) called the moving word task has investigated this ability in children's representations of print. Before learning to read, children believe that the notational forms of print are meaningful objects that function like pictures to indicate words: *d* can be used to mean *dog* and *d d d* can be used to mean *three dogs* (Ferreiro, 1984). Children's error is in representing the letters as objects instead of symbols. Reading requires children to move beyond this object-based understanding to a symbolic representation: *d* stands for a particular sound.

The moving-word task assesses children's understanding of this principle. Children who can recognize letters and print their own names are shown a card with a simple word on it and told what the card says, for example, *dog*. The card is placed under a picture of a dog, accidentally moved to be under a different picture, and then returned to its original place. At each point, children are asked to say what is written on the card. If they understand that the letters on the card function as symbols, they will know that the word always says the same thing because the letters have not changed. Before children have acquired this insight, print notations take on any meaning children ascribe to them. Bialystok (1995), for example, demonstrated that children who can recognize letters but have not yet understood their symbolic function produce cursive-like notations in imitation of adult writing and insist that the writing is meaningful and that those meanings can change. In the moving-word task, children less than about 5 years old change their answer about the identity of the printed word when the card has been moved to align with a different picture.

The errors on the moving-word task are not caused by memory limitations. First, the third time the question is asked, children's memory appears to be restored when the card is again adjacent to the named pictures, making memory lapse an unlikely explanation for errors in the second question. Second, in a manipulation of the moving-word task (Bialystok & Martin, 2003, Study 3), the word printed on the card did not name either of the pictures. In this case, weak memory would produce the usual excellent performance on the first question, then a continuing decline through the second and third questions as children's memory fades. Instead, the results showed that 4-year-old children performed as poorly on the first question when the word did not name the picture as they normally do on the second question after the irrelevant move. Therefore, the results are attributable to the misleading correspondence with the adjacent picture and not to memory failure. It is imperative that children cannot read or the task would be trivially easy, but it is equally imperative that they can recognize letters or the results would be uninterpretable. The representational challenge is to encode the letter–sound relation as the symbolic basis for determining what the word says.

In addition to representational ability, the task requires inhibitory control. When the card is under the wrong picture, children must inhibit the usual tendency to seek meaning from context. Children's experiences with picture books support the strategy of using the graphic information to help assemble meaning; in picture books, text and pictures are normally congruent. In past research, bilingual children have been shown to solve this problem earlier than comparable monolingual children (Bialystok, 1997; Bialystok, Shenfield, & Codd, 2000), an advantage attributed to the enhanced inhibitory control that comes with bilingualism (Bialystok, 2001).

A recent study examined the cognitive basis of the moving-word task by administering the day–night Stroop task (Gerstadt et al., 1994) as a measure of inhibitory control and a concept-learning task as a measure of symbolic representation (Bialystok & Martin, 2003). In the concept-learning task, children learned a relation between a notation (Chinese character) and a meaning and had to apply that meaning to the character when it was encountered later. Scores on the moving-word task correlated with both the concept-learning task, $r(29) = .62$, $p = .0003$, and day-night Stroop task, $r(29) = .48$, $p = .008$, but hierarchical regression analyses showed that only the concept-learning task accounted for unique variance (about 20%) in the moving-word task. Therefore, the moving-word task is associated with measures of inhibition and with measures of representation, but only the representational score uniquely predicts performance. The central cognitive ability in solving the moving-word task, therefore, is

representational ability even though inhibitory control is also involved.

There are two roles for representation in A–R tasks: The first is the ability to represent objects as functional entities and the second is the ability to represent features as identifying attributes of those objects. Both are rooted in symbolic relations in which features signal meanings and remain invariant over time and transformations. For representational objects, the features support both the actual and symbolic identity; therefore, the distinction between reality and appearance questions should be blurred. For real objects, the conflict between appearance and function is more distinct; therefore, these objects should offer a clearer test of the distinction between question types.

Method

Participants

The study included 67 children (39 boys, 28 girls) who were 3, 4, and 5 years old. There were thirty 3-year-olds ($M = 43$ months, $SD = 2.6$), twenty 4-year-olds ($M = 53$ months, $SD = 3.4$), and seventeen 5-year-olds ($M = 63$ months, $SD = 2.9$). All of the children attended nursery school programs and were tested individually in a quiet place in the school. A total of 71 children were tested, but 4 children were not included in the final analysis. One child was absent for an extended period and the other 3 children could not pass the training on most of the executive function tasks.

Procedure and Instruments

There were two testing sessions, each taking approximately 15 min and separated by about 1 week. Six tasks were presented in a fixed order, three in each of the two testing sessions. The first session included forward digit span, day–night Stroop task, and the A–R tasks; the second consisted of Luria's (1996) hand-tapping task, moving-word task, and opposite-worlds task. A stuffed toy known as Tigger was used to explain instructions and facilitate the tasks.

Executive Functioning Tasks

Forward digit span. This task, taken from the Weschler Intelligence Scale for Children–Revised (WISC–R; Weschler, 1974), served as a test of verbal short-term memory. The experimenter read aloud a series of number sequences, ranging in length from two to nine digits, and the child repeated the numbers in the same order. There were two trials for each sequence length. The test ended when children made errors on both trials of a given length. The child's digit span score is the number of trials completed correctly.

Forward digit span is not an executive functioning task because it does not require any operations to be performed on the remembered sequence. In contrast, backward digit span does indicate executive functioning through working memory, but young children find this test to be difficult and many children of the age examined in the present study cannot perform the required transformation. Therefore, the intention was only to obtain short-term memory scores to ensure that children were not at a disadvantage because of memory limitations.

Day–night Stroop task (Gerstadt et al., 1994). This task requires remembering two rules and inhibiting a natural verbal response to a visual stimulus. This task is an adaptation of the original day–night tasked used in Gerstadt et al., (1994). The children were told a story about Tigger bouncing around and ending up in a strange land where all of the names of things are mixed up. They were shown a set of cards in which each card depicted either a blue sky with sun and a tree or a black sky with stars and moon. The experimenter pointed to the night card and said, "When you see this card, I want you to say 'day,'" and the child repeated the word *day*. The experimenter then showed the day card and said, "When you see this card I want you to say 'night,'" and the child repeated the word *night*. These instructions were repeated twice, and children were then given two practice trials. Following this, they were told that they would see more cards one at a time and that they were to name them as fast as possible without making mistakes. The dependent variable was the number of cards named correctly in 30 s. There were 26 cards in the set, yielding a maximum score of 26.

Opposite-worlds task. This task was a modified version of the opposite worlds subtask from the Test of Everyday Attention for Children (TEACh; Manly, Robertson, Anderson, & Nimmo-Smith, 1999). Children need to remember two rules and inhibit their habitual verbal response to a visual stimulus. There were two conditions: In the same-world condition, children named pictures of cows and pigs; in the opposite-world condition, children applied the reverse name to each picture. Sixteen stickers of cows and pigs were arranged in a trail on a blue board, and children were asked to follow the trail and name the animals as quickly as possible without making mistakes. A picture of a barn that was right side up (same world) or upside down (opposite world)

served as a reminder for the rule. The responses were tape recorded and scored for reaction time and accuracy. There were two trials for each of the same- and opposite-world conditions, each producing a score out of 16 for the number of correct responses. The mean of the two trials was the accuracy score for each of the same and opposite conditions. Reaction times for each condition were calculated as the mean of the time to name the set of 16 animals in each of the two trials. The difference score was the mean reaction-time difference between the time needed to name the stickers in the two conditions. The opposite world condition always took longer; therefore, the greater this difference was, the more interfering was the opposite condition for the child. The reaction-time difference score was the measure of inhibitory control in that higher scores indicated weaker inhibition.

Luria's (1966) tapping test. This task was created by Luria (1966) but adapted for experimental use with children by Diamond and Taylor (1996). The task requires the child to remember two rules and to inhibit a motor response to a visual and auditory stimulus. The child was given a "magic wand" and instructed to tap one time after the experimenter tapped twice and to tap twice after the experimenter tapped once. There were 16 trials presented in random order. One point was awarded for each correct trial, creating a maximum score of 16.

Composite inhibitory control. The inhibition tasks were combined into a composite measure of inhibitory control. The tasks had different maximum scores and the interpretation of the reaction-time difference for the opposite-world task was the reverse of the other tasks in that higher scores indicated weaker performance; therefore, an arithmetic sum could not be used. Instead, all the scores were converted to z scores and the z scores for the opposite-worlds task were multiplied by -1 to change the sign from negative to positive, reducing the composite by the amount of that score. Hence, higher scores on this composite indicated better performance in the combined three tasks. The composite score is intended to provide an overall assessment of the various aspects of inhibitory control that are measured separately in the individual tasks even though the individual tasks measure different aspects of inhibitory control.

Moving-word problem. Children were told that Tigger was in a very bouncy mood. The experimenter showed the child two pictures of common objects (e.g., a fish and an umbrella), placed them side by side on the table, and identified them. After this identification, there was no further discussion of the pictures. The experimenter produced a card with the name of one of the pictures (e.g., fish) and told the child what word was written on the card. The card was placed under the named object, in this case the fish, and the child was asked what the card said (introductory question). Suddenly, the experimenter created a situation in which Tigger bounced around the table distracting the child's attention and "accidentally" pushed the card so that it was under the other picture (i.e., the umbrella). The experimenter asked again, "What does this card say?" (incongruent question). Finally, the experimenter commented on the mess and said that it must be tidied up. The card was returned to the original location and the child was asked for the third time what the card said (congruent question). The direction the card was moved was counterbalanced across trials. The change in the display was made to appear as an accident so that the child would not expect that a new answer was necessarily required. Three trials were given using different picture pairs for each (truck and fish, toothbrush and giraffe, flower and glasses) and the words on the cards were truck, toothbrush, and flower, respectively. Children received a score out of 3 for each of the three questions: introductory, incongruent, and congruent.

A–R Tasks

Four items were presented: The real objects were the sponge-rock and the crayon-Lego and the representational objects were the snowman-book and the whale-marker. These items are shown in Figure 1. Three questions were asked about each object following the basic procedure used by Gopnik and Astington (1988). The questions were asked in a fixed order but the objects were presented in counterbalanced order. On each trial, the experimenter put Tigger away for a nap, placed the object on the table, and said, "Look what I have, can you tell me what this is?" Almost all of the children answered correctly (crayons, rock, snowman, whale), many spontaneously labeling the items before the question was asked. The exceptions were 1 child who said "sponge," 1 who said "marker," and several children who called the snowman Frosty. The 2 children who spontaneously labeled the objects as sponge and marker were not included in the analyses for those objects but the children who called the snowman Frosty were included because the concept was correct. Some children needed prompting to name the object. Following this, the object's function or hidden properties were revealed and the following three questions were asked.

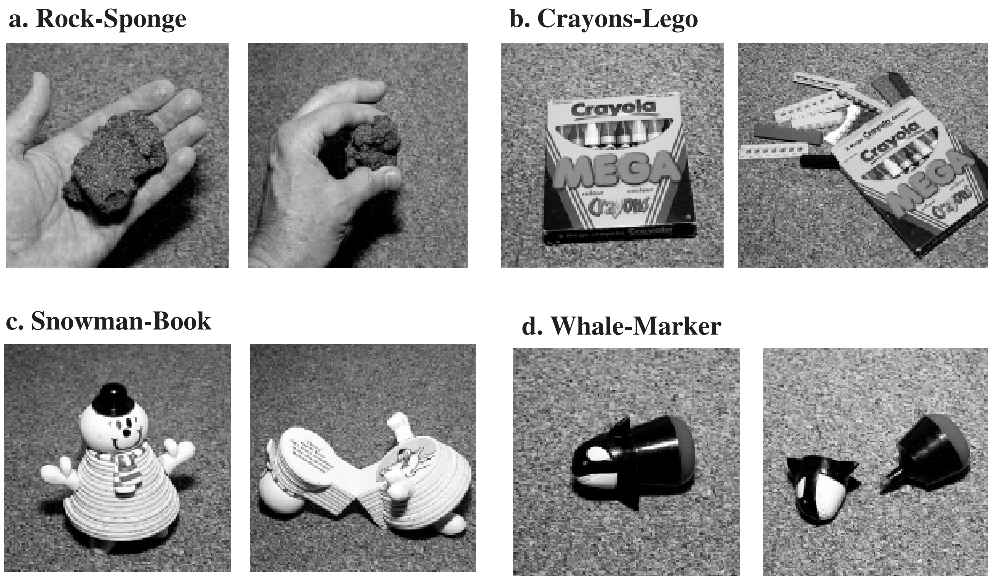

Figure 1. Objects used in appearance – reality task.

Appearance 1: ''What did you think this was when you first saw it?''

Appearance 2: ''Tigger didn't see or hear what we were doing. What will Tigger think this is?''

Reality 3: ''What is it really?''

Questions 1 and 2 are appearance questions because the correct answer requires a description of the object's perceptual features; Question 3 is a reality question because the correct answer is the functional identity of the object. Children who answered all three questions correctly were considered to pass that item. The total A – R score is the proportion of the four items passed by each child.

As explained earlier, the identity of an object is determined by its function. The function is clear for real objects but it is more ambiguous for representational objects because they could serve the role specified by the appearance as a toy. Even though children do not think that the items are actually a snowman and a whale, they could engage in play with them as a toy snowman and a toy whale. Our criterion for ''really'' was confined to actual functions. The book could be a toy snowman, but it could really be read; the marker conveyed the shape and coloring of a whale, but it could really be used to write. Children can reliably distinguish between play and actual functions from 2 1/2 years old, when they begin to engage in symbolic play and accept symbolic functions for common objects (Harris & Kavanaugh, 1993). Therefore, the correct answer to the reality question for the representational items was determined by their functions. Hence, the criteria were consistent across all items: The perceptual features determined the answer to the appearance question and the functional possibilities determined the answer to the reality question.

Results

Executive Functioning Tasks

The mean scores and standard deviations for the executive functioning tasks are reported in Table 1. A one-way analysis of variance (ANOVA) for age on scores for the forward digit span showed no differences among the three groups, $F(2, 64) = 2.18$, *ns*. Digit span scores correlated with the A – R total score, $r(67) = .35$, $p = .003$, and the reality question score, $r(67) = .35$, $p = .003$, but not with the appearance question score, $r(67) = .21$, *ns*. It was also correlated with both executive function measures, specifically, the composite inhibition score, $r(67) = .43$, $p = .0003$, and the moving word score, $r(67) = .25$, $p = .04$. Because of the shared variance between digit span and other measures, digit span was partialled out of the other correlations to control for individual differences in working memory.

A one-way ANOVA for age on the scores for the day – night Stroop task revealed a significant effect, $F(2, 64) = 7.05$, $MSE = 14.63$, $p = .002$, shown by Scheffé comparisons to be the superior performance of the 5-year-olds compared with the other two age groups.

The mean number correct out of 16 in each condition of the opposite-worlds task is reported in Table 1. Although all scores were high, the same-world

Table 1
Mean Scores (Standard Deviations) on Executive Functioning Tasks by Age in Study 1

Age	Digit span	Day-night	Opposite worlds Same	Opposite worlds Opposite	Opposite worlds RT difference	Tapping	Moving word
Maximum		26	16	16		16	3
3 years	6.7 (1.8)	10.5 (3.5)	14.1 (1.6)	12.6 (2.1)	12.6 (10.0)	10.5 (4.0)	1.7 (1.4)
4 years	7.4 (1.7)	11.1 (3.3)	14.7 (1.1)	13.5 (1.9)	7.7 (4.9)	14.1 (2.6)	2.3 (1.1)
5 years	7.8 (2.4)	14.7 (4.8)	15.1 (0.9)	14.8 (1.2)	4.2 (4.1)	14.3 (3.9)	2.9 (0.2)
M	7.2 (2.0)	11.7 (4.2)	14.5 (1.2)	13.4 (1.7)	9.0 (8.2)	12.5 (4.0)	2.2 (1.2)

Note. RT = reaction time.

condition was solved better than the opposite-world condition, $F(1, 64) = 18.87$, $MSE = 1.73$, $p = .0001$, and the 5-year-olds were more successful than the 3-year-olds, $F(2, 64) = 7.65$, $MSE = 3.58$, $p = .001$. The reaction-time difference between same- and opposite-world conditions is also reported in Table 1. A one-way ANOVA for age on these scores showed a significant effect, $F(2, 64) = 7.04$, $MSE = 56.9$, $p = .001$, and Scheffé comparisons revealed the difference to be the contrast between scores for the 5-year-olds and the 3-year-olds, with 4-year-olds not different from either group.

The tapping task scores out of 16 are reported in Table 1. A one-way ANOVA revealed an effect of age, $F(2, 64) = 8.34$, $MSE = 13.29$, $p = .0006$, and Scheffé contrasts showed that both 4- and 5-year-olds performed better than 3-year-olds.

In the moving-word task, all children scored perfectly on the introductory question ($M = 3.0$) and nearly perfectly on the congruent questions (5-year-olds: $M = 3.0$; 3- and 4-year-olds: $M = 2.9$). The important data are the responses to the incongruent question, and these scores are reported in Table 1. A one-way ANOVA for age showed the difference to be significant, $F(2, 64) = 6.05$, $MSE = 0.15$, $p = .004$, and Scheffé contrasts indicated that the significant gap was the difference between the 5- and 3-year-olds.

A–R Tasks

Following the procedure used in most research of this type, children were designated as passing or failing based on their correct responses to both kinds of questions for each item. Children passed an item if they were correct on all three questions. The mean proportion of the four items passed by children is the total proportion score. These scores are reported in Table 2. This success rate is consistent with that reported by Wellman et al. (2001) in their meta-analysis for children of these ages solving similar problems.

A one-way ANOVA examining differences among the three questions was conducted for each of the four objects. All four analyses produced the same pattern: There was an effect of question for each object: sponge-rock, $F(2, 132) = 11.54$, $p = .01$; crayons-Lego, $F(2, 132) = 5.96$, $p = .01$; snowman-book, $F(2, 132) = 5.04$, $p = .01$; and whale-marker, $F(2, 132) = 4.45$, $p = .01$. Contrasts ($p = .05$) indicated no difference between Questions 1 and 2, the two appearance questions, but reliable differences between each of these and Question 3, the reality question. Therefore, the two appearance questions were combined into a single score created by the mean of the two appearance questions for all subsequent analyses. Wellman et al. (2001) also reported that children's responses to questions about self versus other do not differ.

The relation between the two question types and the four objects was examined in a 2 (question) × 4 (object) × 3 (age) ANOVA. The mean scores are reported in Table 3. There was an effect of age, $F(2, 64) = 10.19$, $MSE = 0.3$, $p = .0001$, indicating a difference between 5-year-olds and 3-year-olds that did not interact with the other variables. There was an

Table 2
Proportion (Standard Deviation) of Children Passing the Appearance–Reality Task by Age and Item in Study 1

	Sponge-rock	Crayon-Lego	Snowman-book	Whale-marker	Proportion passed
3 years	.13 (.32)	.26 (.43)	.23 (.41)	.30 (.43)	.23 (.32)
4 years	.40 (.51)	.55 (.52)	.55 (.54)	.60 (.50)	.53 (.31)
5 years	.53 (.50)	.76 (.42)	.58 (.53)	.58 (.52)	.61 (.33)

Table 3
Mean Scores (Standard Deviations) on the Theory-of-Mind Tasks by Object and Question

Object type	Object	Question	3 years	4 years	5 years
Real	Sponge-	Appearance –	0.37 (.39)	0.57 (.40)	0.67 (.39)
	rock	reality	0.67 (.48)	0.80 (.41)	0.94 (.24)
	Crayon-	Appearance –	0.53 (.45)	0.72 (.37)	0.82 (.35)
	Lego	reality	0.79 (.42)	0.95 (.22)	1.00 (.00)
Representational	Snowman-	Appearance –	0.72 (.40)	0.88 (.22)	0.83 (.30)
	book	reality	0.47 (.51)	0.65 (.49)	0.82 (.39)
	Whale-	Appearance –	0.80 (.36)	0.92 (.18)	0.80 (.35)
	marker	reality	0.50 (.51)	0.75 (.44)	0.82 (.39)

effect for object, $F(3, 192) = 3.67$, $MSE = 0.3$, $p = .01$, showing that the sponge-rock was more difficult than the other three items. Finally, there was an interaction of object and question, $F(3, 192) = 13.17$, MSE = 0.1, $p = .0001$. Simple effects analyses ($p = .01$) indicated that children scored higher on the appearance questions than on the reality questions for whale-marker and snowman-book but they scored lower on the appearance questions than on the reality questions for crayon-Lego and sponge-rock. This grouping supports the classification of the two object types. The interaction is plotted in Figure 2.

A different way of considering the interaction between object type and question type is to examine the relation between scores on these variables. A correlational analysis revealed a relation between children's ability to answer the same question type across objects, but no relation between the two different object types. Both appearance questions,

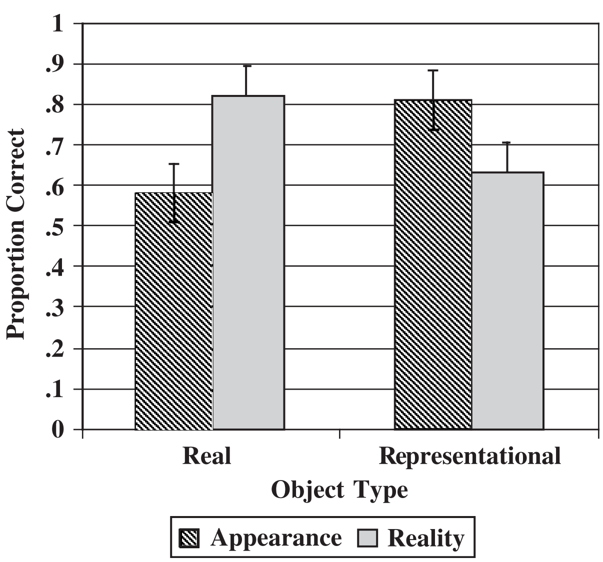

Figure 2. Mean scores for theory-of-mind tasks; interaction of object and question in Study 1.

$r = .48$ $p = .01$, and reality questions, $r = .53$, $p = .004$, correlated across real and representational objects, but the correlations between object types was negligible ($r = .05$ and $r = -.09$ for real and representational objects, respectively). Therefore, there is more common processing involved in answering the same type of questions than in operating on the same type of objects.

The relation between performance on the executive functioning tasks and success in the A – R tasks was examined in a series of hierarchical regression analyses using the total A – R score for the proportion of items passed, the appearance question score, and the reality question score as dependent variables. Similar models assessed the two object types as dependent variables, but because there were no significant patterns they are not reported. The correlation pattern reported earlier suggests that the reason is that there is more systematic variance associated with question type than with object type. The correlation matrix for the variables in the regression analyses is reported in Table 4. These correlations were calculated by partialling out the variance from digit span to control for the contribution of working memory to these scores. The

Table 4
Partial Correlation Among Executive Function Variables and Appearance – Reality (A – R) Scores Controlling for Digit Span

	1	2	3	4	5	6
1. A – R total						
2. Appearance question	.75*					
3. Reality question	.52*	.04				
4. Moving word	.33*	.30*	.19			
5. Day-night	.12	– .05	.31*	.15		
6. Cow-pig	– .30*	– .16	– .32*	– .29*	– .30*	
7. Tapping	.26*	.08	.28*	.35*	.17	– .05

*$p < .01$.

Table 5
Regression Analysis on Appearance – Reality Task

Variable	Parameter	*SD* error	ΔR^2	*t*	*p*
Total score: $F(3, 63) = 7.57$, $MSE = 0.09$, $p = .0001$, $R^2 = .33$					
Age	.01	.01	.25	2.58	.01
Digit span	.03	.02	.13	1.70	*ns*
Inhibition	.01	.02	.13	0.42	*ns*
Moving word	.05	.03	.14	1.61	*ns*
Appearance question: $F(3, 63) = 3.38$, $MSE = 1.16$, $p = .02$, $R^2 = .17$					
Age	.02	.02	.07	1.40	*ns*
Digit span	.09	.07	.02	1.33	*ns*
Inhibition	− .02	.02	.01	− 1.10	*ns*
Moving word	.79	.36	.07	2.29	.03
Reality question: $F(3, 63) = 7.41$, $MSE = 1.02$, $p = .0001$, $R^2 = .32$					
Age	.03	.02	.22	0.57	*ns*
Digit span	.08	.07	.05	1.28	*ns*
Moving word	− .01	.11	.00	− 0.1	*ns*
Inhibition	.17	.08	.05	2.26	.02

main pattern in these correlations is that the appearance question is correlated only with the moving-word task and the reality question is correlated with all three inhibition tasks but not the moving-word task. The table includes the three inhibition measures separately, although the composite score was used in the regression analyses.

All three models were constructed by entering age and digit span in the first two steps to establish the contribution of the basic developmental variables, followed by the composite inhibition score and the moving word score, the last two being entered in both orders to isolate the unique variance associated with each. The results of these models are reported in Table 5. The model using the total score as the dependent variable accounted for 33% of the variance, but only age was significant in predicting this overall performance score. Using scores on the appearance questions as the dependent variable, the model accounted for 17% of the variance, and the only significant contribution was made by scores on the moving-word task. Using scores on the reality questions as the dependent variable, the model accounted for 32% of the variance, and the only significant predictor was the measure of inhibitory control.

Discussion

Consistent with other research, children improved in their ability to solve A – R problems between the ages of 3 and 5 years, although by 5 years of age the children were still not performing at ceiling. This finding is compatible with theory-theory interpretations of this emerging ability, and the overall success rate was comparable to that reported in the literature for children of this age. Also consistent with other research, there were significant correlations between children's success on the A – R task and their performance on other cognitive measures, including short-term memory, inhibition, and representation. This finding is compatible with executive function interpretations of these problems.

The main contribution of the present study is in the detail it provides about the role of specific executive (and cognitive) functions in aspects of the A – R task. Children's success depended on the type of object being described and the type of question being asked. For real objects, reality questions were more easily solved than appearance questions, and for representational objects the reverse was true. More important, children's ability to answer each of the question types was determined by different cognitive factors: Success on the appearance questions, irrespective of object type, was related to scores on the moving-word problem, considered to be an indication of representational ability, and success on the reality question was related to scores on the inhibition problems.

The difference between question types was more influential than the difference between object types in determining children's solutions as shown by the pattern of correlations. The main difference between the object types was in the degree of the conflict between their perceptual appearance and their functional utility. The main difference between the question types was in the cognitive processes recruited to answer each. Thus, the interaction be-

tween these factors showed that some types of items are more amenable to some types of processing, but the pervasive finding, confirmed both by ANOVA and the regression models, is that the cognitive processes change with the question types.

All the tasks required children to hold a description in mind in the presence of conflicting information; for the A–R task, the misleading information is the conflicting appearance of the object, for the moving word task, it is the irrelevant picture, and for the inhibition tasks it is an image or model that is the opposite of the required response. Thus, all these tasks require both the ability to represent and recall information and the ability to control attention to that information in a conflicting situation. In the terms used by Diamond (2002), the tasks require both working memory and inhibition. However, the bias that makes one or the other of these components more central to the solution of each task is different.

For the moving-word task, the central process is the ability to construct the representation that relates the object and meaning, a representation that is inherently symbolic (stands for) and developmentally later than representations that identify objects directly through their perceptual features (is a; Bialystok & Martin, 2003). This is similar to the ability needed to respond to the appearance questions. In the appearance question, the perceptual features are consistent with the correct solution—the item looks like a rock, even though contradictory information about its identity has been introduced. The question requires an accurate representation of an earlier situation: When I first saw this, I thought it was a rock. This representation is the basis for answering both questions. This is similar to the problem posed in the moving-word task. The child is told that the card says "fish," and then a spatial transposition introduces conflicting information from an irrelevant picture. When the child is asked again what the card says, the child must rely on a clear representation of that meaning: When I first saw this card I was told it said "fish," even though it is tempting to think about its relation to the new picture. Thus, the appearance question and the incongruent question in the moving-word task rely on the stability of the representation. In both cases, that stability is challenged by new information: the object's functional identity in the A–R task and the adjacency to an irrelevant picture in the moving-word task.

For the inhibitory control tasks, the central process is the ability to ignore the information that is opposite to that required for the response, for example, a picture of day when the response must be "night." This is a more demanding challenge than that posed by irrelevant information; the information in this case is automatically invoked by its presentation, just as the meaning of the color word is in a Stroop task. This is similar to the ability needed to respond to the reality questions. To answer the reality question, children must ignore the fact that the item looks like a rock and trust the new information that it is actually a sponge. In this case, the perceptual features lead to an opposite and incorrect response. Even though the object continues to look like a rock, the child must accept that it is a sponge. The reality questions are easier for the real objects. The reason may be in the centrality of functional information to object identity, especially for children. Having learned that the object is a sponge and can only function as a sponge, it ceases to be a rock. Representational objects do not present a clear divide between appearance and function; having learned that the object is a book, it can still function as a toy snowman. For these materials, there is no contradiction between being both a (toy) snowman and a book, or both a (toy) whale and a marker. It is less likely that a rock can function as a sponge; it is unexpected (but not impossible) for a box to contain both crayons and Legos (assuming only one type of contents). Therefore, the reality question is more ambiguous for representational objects than for real objects. For these objects, it is the appearance and the perceptual features that are most salient, and these features support the correct solution to the appearance question. Therefore, there is a significantly larger difference between the appearance and reality questions for the real objects than there is for the representational objects.

This account depends on a conception of processing in the solution to these questions that distinguishes between the cognitive components involved in representation and inhibition. The regression model using total score as the dependent variable confirmed the age-related improvement in this problem that is generally reported in the literature. The more detailed analyses of the question types, however, revealed different processes and developmental trajectories for the components of the problem. This view was investigated further in Study 2 by comparing groups of children who differ in their initial abilities to inhibit attention to misleading information.

Study 2

In Study 1, the ability to answer the two question types was related to different executive processes:

Representation was related to the appearance questions and inhibition was related to the reality question. Children who have developed representation and inhibitory control to different levels, therefore, may provide another means of demonstrating a distinction between the processes involved in these two question types.

In previous research we have shown that bilingual preschoolers are more advanced than their monolingual peers on measures of inhibitory control (summarized in Bialystok, 2001). The research has examined children performing tasks in different domains, usually by comparing performance on pairs of tasks that are similar in representational demands but differ in their need for inhibitory control. A metalinguistic task, grammaticality judgment, can be manipulated by training children to judge the acceptability of sentences on the basis of their structure ("Apples grow on trees") but then asking them to judge sentences that are either ungrammatical ("Apples trees on grow") or anomalous ("Apples grow on noses"). The ungrammatical sentences require a representation of correct structure to detect the error, but the anomalous sentences require inhibitory control to ignore the misleading anomaly. Repeated studies have shown that monolinguals and bilinguals perform the same on ungrammatical items but bilinguals surpass monolinguals on anomalous items (Bialystok, 1986, 1988; Cromdal, 1999). A study based on numerical concepts of cardinality has shown the same pattern: A task assessing children's knowledge of equivalent set size when the numerical quantity is the same was solved equally by monolinguals and bilinguals, but a task placing the set size judgment in a misleading context where the smaller set created a higher tower was solved better by bilinguals (Bialystok & Codd, 1997). Finally, nonverbal problems yield the same dissociation. The Noelting juice task, an analytic measure of knowledge of proportions without misleading information, was solved equally by monolinguals and bilinguals, but the water-level task, in which misleading information from the position of a tilted beaker needs to be ignored, was solved better by bilinguals (Bialystok & Majumder, 1998). Therefore, bilingual children consistently surpass monolingual children on tasks requiring inhibitory control but perform the same as monolinguals on tasks based more on representational processes without misleading information. Therefore, bilingual children should outperform monolinguals on solutions to the reality question. There is no reason to expect performance differences between the two groups in solutions to the appearance questions. These predictions were tested in Study 2.

Method

Participants

The study included 95 children (48 boys, 47 girls) who were 4 to 5 years old. There were 52 monolinguals ($M = 57$ months, $SD = 8.6$) and 43 bilinguals ($M = 58$ months, $SD = 9.7$), divided into two age groups, described as 4-year-olds (33 monolinguals: $M = 52$ months, $SD = 4.4$; 22 bilinguals: $M = 51$ months, $SD = 5.7$) and 5-year-olds (19 monolinguals: $M = 67$ months, $SD = 4.2$; 21 bilinguals: $M = 67$ months, $SD = 4.5$).

The bilingual children lived in an area populated by many immigrant families and represented a wide range of language backgrounds. All of the children spoke their first language at home with their family and used English at school and in the community. The languages spoken by the bilingual children included: Armenian, Filipino, Hebrew, Russian, Tagalog, Spanish, Arabic, African, Chinese, Somali, Hungarian, Polish, Persian, Mandarin, and French. There was a tendency for the monolingual children to live in more middle-class neighborhoods than those of the bilinguals, suggesting that the monolingual children were higher in socioeconomic status. This difference, however, makes the prediction for a bilingual advantage a more stringent test of the hypothesis. All the children were recruited from their schools and were tested in a quiet place in the school.

Tasks and Procedures

Children were tested individually and received all three tasks in one session. The order of presentation of the tasks was forward digit span, Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981), and the four A–R tasks.

PPVT–R. All children were given the PPVT–R in English. It was not possible to test the bilingual children in their home language because of the diversity of these languages. However, English was the second language for all of these children because they spoke their first language at home with their families; therefore, their proficiency in English provided a measure of their bilingualism. In addition, it was important to quantify their English skills because the task was presented only in English.

Digit span. Digit span was assessed using a different procedure from that followed in Study 1. In the present study, the rules were explained and testing began with at string of two digits, just as in Study 1. Unlike Study 1, however, if children re-

sponded correctly, the experimenter moved on to a string of the next length, in this case, three digits. If a child made an error, a second string of the same length was offered. If this was correct, the experimenter moved on to the next length. When both strings of the same length were wrong, testing stopped and the digit span score was the longest string length in which children successfully responded to one of the items. The digit span scores, therefore, are necessarily lower than those recorded in Study 1 because children typically receive only one trial for each length.

A–R task. As in Study 1, there were four objects classified as real or representational. The two real objects were the same as those used in Study 1, namely, the sponge-rock and the crayons box containing Legos. The representational objects included the snowman-book and a new item. A shallow white bowl, 5 in. in diameter, was marked on the inside bottom with a red happy face. Looking into the bowl, the object was a picture of a happy face. When a red cellophane film was placed over the bowl, the face became invisible and the bowl appeared empty. Therefore, the item changed from a picture of a happy face to an empty bowl. Although it would have facilitated comparison across the two studies if the procedures had been identical, the studies were conducted independently.

Because there was no difference between the two appearance questions in Study 1, the two questions were counterbalanced across children and objects in this study. For half of the objects (one real and one representational), half of the children were asked what they thought the object was when they first saw it and the other half were asked what Tigger would say it was. For the other half of the objects, the opposite question was asked.

Results

The 4-year-old ($M = 100.7$, $SD = 26.6$) and 5-year-old ($M = 98.8$, $SD = 19.9$) monolinguals scored higher than the 4-year-old ($M = 82.1$, $SD = 22.7$) and 5-year-old ($M = 83.0$, $SD = 16.9$) bilinguals on the PPVT–R, $F(1, 91) = 13.97$, $MSE = 509.6$, $p = .0003$. There was no age difference and no interaction. In contrast, the 5-year-old monolinguals ($M = 4.6$, $SD = 1.5$) and bilinguals ($M = 4.7$, $SD = 1.2$) scored higher than the 4-year-old monolinguals ($M = 3.8$, $SD = 1.4$) and bilinguals ($M = 3.7$, $SD = 1.4$) on the digit span scores, $F(1, 91) = 8.65$, $MSE = 1.9$, $p = .004$. There was no language group difference and no interaction. Because a different method for presentation and scoring was used from that employed in Study 1, the results cannot be directly compared. Nonetheless, the digit span score in both cases provides an approximation for short-term verbal memory even though the two methods place those results on a different scale.

The relation between scores on the background measures of receptive vocabulary and digit span and the two question types in the A–R task was evaluated. Informally, the linguistic structure of the appearance questions is more complex than that of the reality questions; therefore, it may be that linguistic ability has a biased influence on the former, confounding the correlation between the two measures. Furthermore, it may be that a relation between linguistic ability and appearance questions was the underlying cause of the correlation between the moving-word task and the appearance questions in Study 1. Therefore, it is important to show that linguistic ability is not specific to appearance questions. However, the results showed that the PPVT–R standard scores were related to the overall proportion of items passed, $r(90) = .47$, $p = .0001$, but were additionally related to scores on both appearance questions, $r(90) = .38$, $p = .0002$, and reality questions, $r(90) = .42$, $p = .0001$. Digit span was also correlated with the overall proportion of items passed, $r(90) = .26$, $p = .01$, but in contrast to vocabulary, it was related only to scores on the reality questions, $r(90) = .32$, $p = .002$, and not the appearance questions, $r(90) = .17$, *ns*. These correlations replicate the results of Study 1 and confirm the dissociation between the cognitive processes involved in these questions reported in that experiment.

As in Study 1, children were classified as passing each item if they answered both questions correctly. The overall proportion of items passed in the A–R task was the mean of these four pass–fail scores. The proportion of children passing each of the four items and the overall proportion of items passed by age and language group are reported in Table 6. Although the mean proportion of children passing in each of the age and language groups appears different, a three-way ANOVA for age, language group, and object type indicates a significant effect only for object, $F(3, 81) = 5.33$, $MSE = .07$, $p = .001$, and no effect of either grouping factor and no interaction. Contrasts indicated that sponge-rock and the happy face were both more difficult than the other two objects (but not different from each other). In Study 1, only the sponge-rock was more difficult than the other three items. The overall passing rate was lower in this study than it was in Study 1. Direct comparisons between performance in the two studies is impossible, but anecdotally, the children in Study 1

Table 6
Proportion (Standard Deviation) of Children Passing the Appearance–Reality Task by Age and Item in Study 2

Age	Language	Sponge-rock	Crayon-Lego	Snowman-book	Happy face	Proportion passed
4 years	Monolingual	.42 (.54)	.44 (.52)	.45 (.53)	.21 (.39)	.33 (.31)
	Bilingual	.37 (.43)	.26 (.44)	.42 (.49)	.12 (.30)	.27 (.23)
5 years	Monolingual	.53 (.51)	.58 (.53)	.42 (.51)	.42 (.52)	.48 (.32)
	Bilingual	.38 (.48)	.54 (.50)	.58 (.52)	.22 (.38)	.40 (.32)

were from more enriched home and school environments.

The responses to the A–R questions were used as variables in a four-way ANOVA for object type, question type, age, and language group. Older children obtained higher scores than younger children, $F(1, 83) = 4.74$, $MSE = 0.2$, $p = .03$, but there was no difference between the two language groups and no interaction (both $F < 1$). For the within-subjects factors, there was an effect of question type, $F(1, 86) = 36.13$, $MSE = 0.1$, $p = .0001$, and an interaction of object type and question type, $F(1, 86) = 4.62$, $MSE = 0.1$, $p = .03$. Simple effects analysis indicated that children obtained higher scores on the reality questions than on the appearance questions for the real objects, $F(1, 91) = 32.89$, $MSE = 0.1$, $p = .0001$, but that both the reality questions and the appearance questions were equivalent for the representational objects. The interaction is plotted in Figure 3. Unlike Study 1, the difference between the questions for the representational objects was not significant but the results replicate the finding in Study 1 that there was a larger difference between question types for real objects than for representational objects. To examine the difference more carefully, an analysis was conducted separately for the two representational objects. The snowman-book produced the pattern from Study 1 in which the appearance questions were easier than the reality questions, $t(89) = 6.26$, $p = .0001$, but the happy face yielded equivalent performance for both question types.

Research in ToM has revealed the importance of language proficiency in these abilities (Astington, 2000; Astington & Jenkins, 1999; de Villiers, 1999; de Villiers & de Villiers, 2000; Happé, 1995; Plaut & Karmiloff-Smith, 1993). Consistent with previous research, the bilingual children in the present study were disadvantaged relative to the monolinguals in a formal test of language proficiency (see Oller & Eilers, 2002, for a review). The attempt to determine whether bilingual children are responding differently to A–R problems, therefore, may be masked by group differences in language proficiency. To compensate for differences in language skills, performance on the A–R task was examined by taking account of initial differences in language proficiency by using analysis of covariance (ANCOVA) with PPVT–R standard scores as a covariate.

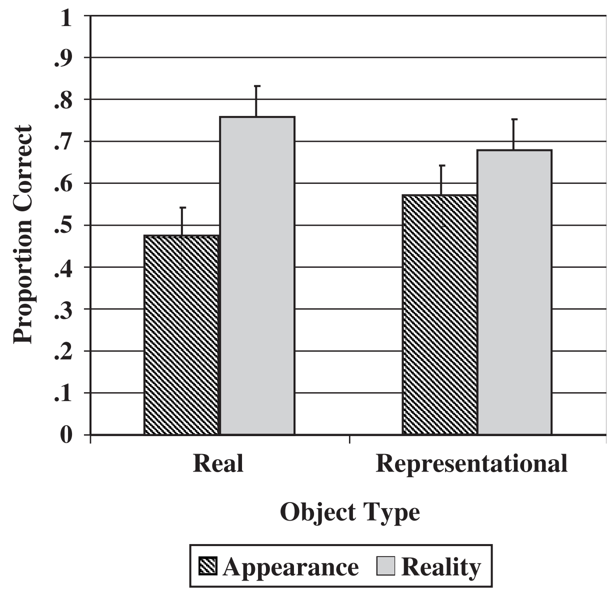

Figure 3. Mean scores for theory-of-mind tasks; interaction of object and question in Study 2.

A 2 (question type) × 2 (object type) × 2 (language group) ANCOVA, using PPVT–R standard scores as the covariate, revealed an effect of question, $F(1, 83) = 4.0$, $MSE = 0.1$, $p = .04$, and an interaction of question and language, $F(1, 83) = 3.64$, $MSE = 0.1$, $p = .05$. For appearance questions, the two language groups performed the same ($F < 1$), but for reality questions, the bilinguals scored reliably better than the monolinguals, $F(1, 91) = 5.81$, $MSE = 0.1$, $p = .01$. The least square means are reported in Table 7.

Discussion

There were two main results. First, the interaction of question type and object type was partly rep-

Table 7
Least Square Means (Standard Deviations) for Object Type and Question Type by Language Group

		Question type	
Language group	Object type	Appearance	Reality
Monolingual	Real	.48 (.06)	.67 (.04)
	Representational	.56 (.04)	.66 (.04)
Bilingual	Real	.48 (.06)	.85 (.05)
	Representational	.59 (.04)	.77 (.05)

licated. As in Study 1, the reality questions were easier than appearance questions for the real objects; unlike Study 1, the reality questions and the appearance questions were equivalent for representational objects. Nonetheless, the results replicated the finding of a greater difference between question types for real objects than for representational objects that was reported in Study 1. The difference between the two studies is probably due to the change in one of the representational objects because the snowman replicated the findings of Study 1. One difference between the happy face and all of the other items is that it is lacking in function. The bowlness that defines the object does not change regardless of whether there is a face at the bottom. Therefore, the object does not present the same kind of conflict as the other items used in these tasks. This partial replication confirms the priority of reality questions for real objects and supports the conclusion that there is a different relation between the two questions for representational objects; even in the absence of a question effect, the reality questions were not easier than appearance questions, as they were for real objects.

Second, the bilingual children performed better than the monolingual children on the reality questions once differences in language facility had been controlled, but the two groups performed the same on the appearance questions. The interpretation of this group difference is that the reality questions require high levels of inhibitory control (Study 1) and bilingual children are more advanced than monolinguals in developing this control (Bialystok, 2001).

In the ANOVA uncorrected for vocabulary, there was no difference between monolinguals and bilinguals in their solutions to the two questions; the difference emerged only when language competence was included in the analysis. Correlation analyses showed that the vocabulary scores were equally correlated with both appearance and reality questions. These data are further evidence for the importance of language proficiency in A–R problems and perhaps ToM problems more generally.

General Discussion

Children solved A–R problems using two kinds of objects that differed in their representational demands and answered two kinds of questions that differed in their cognitive requirements. These two distinctions interacted. Reality questions for representational objects were confusing because the designation of one of the functions as more central is arbitrary: Is it primarily a toy whale or is it primarily a marker? This is not the case for the real objects: The item may resemble a rock but it cannot carry out any functions served by a rock. The appearance questions were easier for the representational objects than for the real objects. For the representational items, the appearance does not strictly contradict the functional identity because the object is perceived to be a toy and can therefore have some other function as well. It is easier for children to contemplate both functions and accept the appearance for the two identities: It is a snowman and it is a book. Therefore, the appearance questions are easier because the more ambiguous function of the representational objects allows children to hold in mind both representations for those objects. Children's representations of objects are driven by their understanding of the function for those objects.

The differences captured by object type and question type provide a means for examining the cognitive basis of the A–R task. The question types invoked different cognitive skills: Appearance questions were related to the moving-word task, primarily a problem of symbolic representation, and reality questions were related to measures of inhibitory control. Other studies have also revealed subtle differences in the cognitive bases of these problems. Hughes (1998a) found that both task type (predict vs. explain) and structure (change in location vs. unexpected contents) affected the relationship among ToM tasks. Hala et al. (2003) reported that change-in-location tasks were correlated and unexpected-contents tasks were correlated, but the two were not related to each other. In the present study, children's performance depended on the type of object being used and the type of question being asked for different items within the A–R paradigm. As in the research by Hala et al., there were correlations in performance across question type but not across object type.

The difference in the underlying cognitive processes required for the two question types was dem-

onstrated by the differential performance between monolingual and bilingual children on these problems. Although we did not give independent tests of inhibitory control to the children in Study 2, past research with similar children has consistently shown the superiority of bilinguals in solving tasks requiring inhibition of attention. For example, all else being equal, bilingual children solve the moving-word task earlier than bilinguals (e.g., Bialystok, 1997); once representational ability is equated, the bilingual advantage in inhibition gives those children the extra boost to solve the task. Consequently, the hypothesis was that bilingual children would also excel in answering the reality questions, shown by the regression analysis in Study 1 to depend on this process. The results confirmed this prediction, but only after language proficiency had been controlled. It is important to note that the bilingual children demonstrated no other cognitive advantage over the monolinguals, and no overall advantage in general success with the A – R task. The advantage was confined to the question based on the executive process developed as a premium benefit of childhood bilingualism. Therefore, we have no reason to believe that bilingual children are more advanced than monolinguals in the conceptual basis of ToM that is captured by the theory theory. Rather, the bilingual advantage allows them to solve typical ToM tasks at a more advanced level than would be expected by their conceptual stage.

The distinction between appearance and reality questions and their interactions with real and representational objects is reminiscent of the distinction between real and phenomenist errors introduced by Flavell et al. (1983) in the original study. Children who reported only the appearance of the object for both questions were committing the phenomenist error, and these occurred mostly when children needed to distinguish between real and apparent perceptual features. Children who reported only the actual properties, ignoring the appearance, committed realist errors, and these occurred when the identity of the stimulus was at issue. Although similar to the distinction proposed here, the realist versus phenomenist categories are largely descriptive. Flavell et al. reported that the two types of errors are associated with different types of tasks. Our attempt was to extend that observation by identifying the underlying processes that make one or the other of these errors more likely.

The representation and inhibition skills examined in these studies are aspects of executive functioning, and to that end, these results contribute to the view that the development of executive function is central to performance for A – R specifically and ToM more generally. In most research investigating children's acquisition of ToM, young children are presented with a series of problems and then are classified as having passed or failed the set based on some performance criterion. This method allows one to determine the point at which children develop the insight necessary to pass these tasks, a point that is often interpreted as the onset of a ToM. The approach in the present studies was different. Our concern was not with identifying the point at which children begin to pass these problems but with the development of the processes that underlie their solutions. Consequently, many of the children in our studies were older than those traditionally used in this research, and most of them were in the transition to being competent solvers of these tasks. The method, however, allowed us to identify the specific cognitive processes involved in these tasks and their relation to children's overall success. As children became more competent with the executive function components involved in these tasks, they also became more successful in solving the A – R problem. The regression analyses demonstrated the specific connections between each of the components of executive function and success in one of the question types.

In the meta-analysis conducted by Wellman et al. (2001), the bulk of the evidence supported the interpretation that there is a holistic shift into understanding ToM. The present results contribute to an explanation of what might underlie that shift. The constellation of skills and knowledge that conspire to guarantee that children can solve these problems indeed may represent a unique arrangement, giving ToM its domain-specific character. Moreover, there is no doubt that when this constellation of factors is in place and children can solve these problems, other significant developments become possible, making ToM itself an explanatory construct in development. But how do children come to solve these problems in the first place? The present studies pointed to the importance of developing ability in aspects of executive functioning and demonstrated how the performance of children who are close to solving ToM problems depends on their progress in these basic cognitive skills. ToM may indeed be a theory, but it does not exist in a vacuum, and its development is part of the larger fabric of the child's growing cognitive competence.

References

Astington, J. W. (2000). Language and metalanguage in children's understanding of mind: Essays in honour of

David Olson. In J. W. Astington (Ed.), *Minds in the making* (pp. 267–284). Oxford, England: Blackwell.

Astington, J. W., & Gopnik, A. (1991a). Developing understanding of desire and intention. In A. Whiten (Ed.), *Natural theories on mind: Evolution, development and stimulation of everyday mindreading* (pp. 39–50). Oxford, England: Blackwell.

Astington, J. W., & Gopnik, A. (1991b). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9*, 7–31.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*, 1311–1320.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*, 498–510.

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24*, 560–567.

Bialystok, E. (1991). Letters, sounds, and symbols: Changes in children's understanding of written language. *Applied Psycholinguistics, 12*, 75–89.

Bialystok, E. (1995). Making concepts of print symbolic: Understanding how writing represents language. *First Language, 15*, 317–338.

Bialystok, E. (1997). Effects of bilingual and biliteracy on children's emerging concepts of print. *Developmental Psychology, 33*, 429–440.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. New York: Cambridge University Press.

Bialystok, E., & Codd, J. (1997). Cardinal limits: Evidence from language awareness and bilingualism for developing concepts of number. *Cognitive Development, 12*, 85–106.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*, 69–85.

Bialystok, E., & Martin, M. M. (2003). Notation to symbol: Development in children's understanding of print. *Journal of Experimental Child Psychology, 86*, 223–243.

Bialystok, E., Shenfield, T., & Codd, J. (2000). Languages, scripts, and the environment: Factors in developing concepts of print. *Developmental Psychology, 36*, 66–76.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology, 18*, 279–295.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish-English bilinguals. *Applied Psycholinguistics, 20*, 1–20.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25–31.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science, 238*, 1556–1557.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development, 62*, 736–752.

DeLoache, J. S., Miller, K. F., & Rosengren, K. S. (1997). The credible shrinking room: Very young children's performance with symbolic and nonsymbolic relations. *Psychological Science, 8*, 308–313.

de Villiers, J. (1999). Language and theory of mind: What are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental neuroscience* (pp. 83–123). New York: Oxford University Press.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determination and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). East Sussex, England: Psychology Press.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York: Oxford University Press.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and "Do as I say, not as I do". *Developmental Psychology, 29*, 315–334.

Duncan, J. (1986). Disorganization of behaviour after frontal lobe damage. *Cognitive Neuropsychology, 3*, 271–290.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test–Revised Form M*. Toronto, Canada: Psycan.

Ferreiro, E. (1984). The underlying logic of literacy development. In H. Goelman, A. Oberg, & F. Smith (Eds.), *Awakening to literacy* (pp. 154–173). Exeter, NH: Heinemann Educational Books.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance–reality distinction. *Cognitive Psychology, 15*, 95–120.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Gerstadt, C. L., Hong, Y.-J., & Diamond, A. (1994). The relationship between cognition and action: Performance

of children 3½–7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153.

Gopnik, A. (1996). Theories and modules: Creation myths, developmental realities, and Neurath's boat. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 169–183). New York: Cambridge University Press.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Development*, *59*, 26–37.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirshfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.

Hala, S., Hug, S., & Henderson, A. (2003). Executive function and false-belief understanding in preschool children: Two tasks are harder than one. *Cognition and Development*, *4*, 275–298.

Hala, S., & Russell, J. (2001). Executive control with strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology*, *80*, 112–141.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, *66*, 843–855.

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society of Research in Child Development*, *58*(Serial No. 231).

Hughes, C. (1998a). Finding your marbles: Does preschoolers' strategic behaviour predict later understanding of mind? *Developmental Psychology*, *34*, 1326–1339.

Hughes, C. (1998b). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233–253.

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Luria, A. R. (1973). *The working brain*. New York: Basic Books.

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *Test of Everyday Attention for Children: TEACh.* Thurston, England: Thames Valley Test Company.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*(Serial No. 136).

Oller, D. K., & Eilers, R. E. (Eds.). (2002). *Language and literacy in bilingual children*. Clevedon, England: Multilingual Matters.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development*, *73*, 752–767.

Plaut, D. C., & Karmiloff-Smith, A. (1993). Representational development and theory-of mind computations. *Behavioral and Brain Sciences*, *16*, 70–71.

Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and explaining behaviour: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, *35*, 1059–1075.

Tager-Flusberg, H., Sullivan, K., & Boshart, J. (1997). Executive functions and performance on false belief tasks. *Developmental Neuropsychology*, *13*, 487–493.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Weschler, D. A. (1974). *Wechsler Intelligence Scale for Children–Revised*. San Antonio, TX: Psychological Corporation.

Whiten, A. (1997). The Machiavellian mindreader. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II* (pp. 144–173). Cambridge, England: Cambridge University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in children's understanding of deception. *Cognition*, *13*, 103–128.